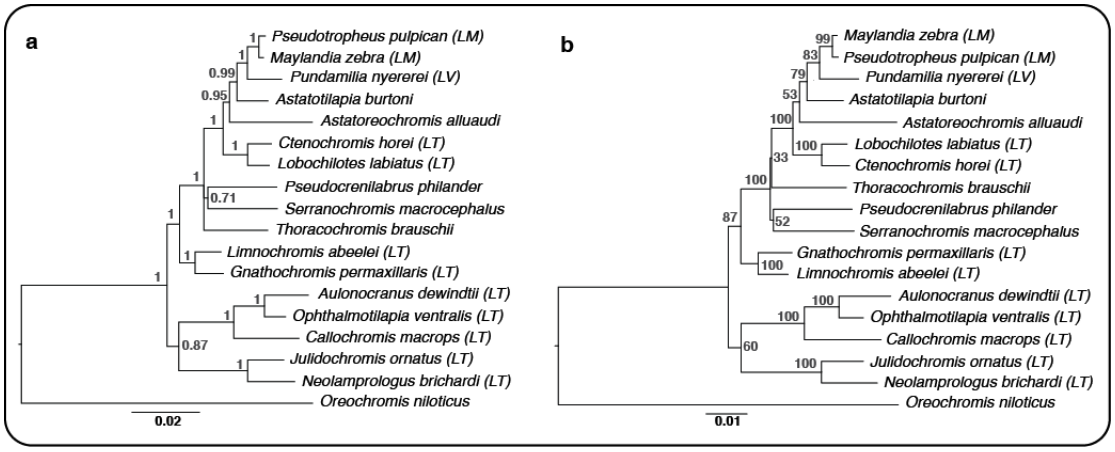
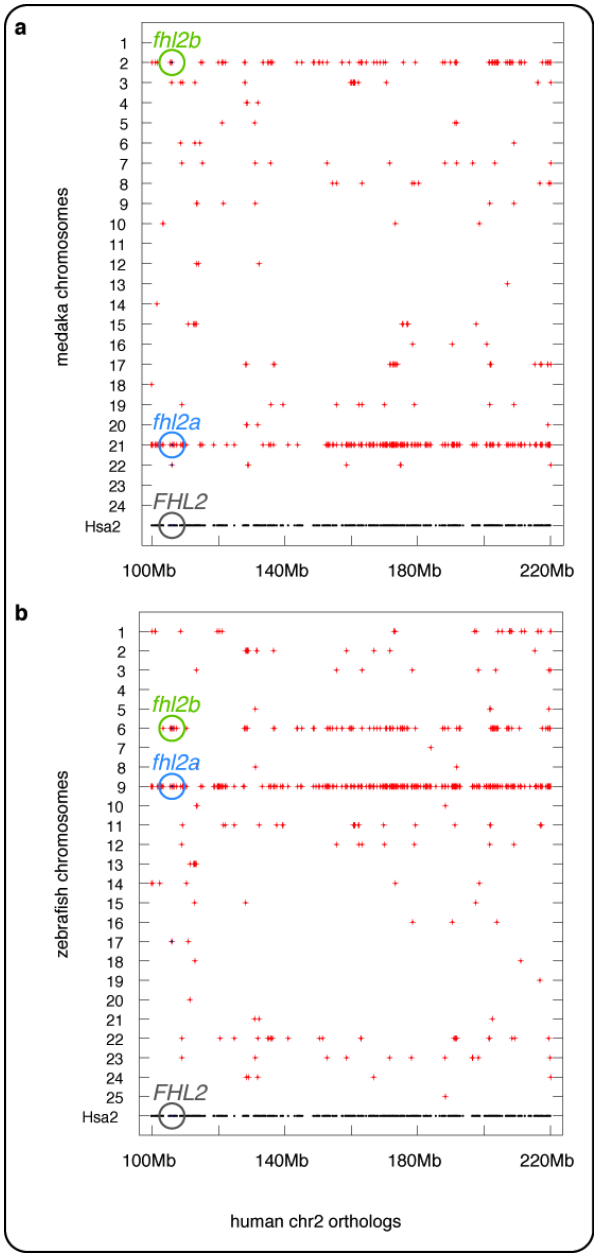
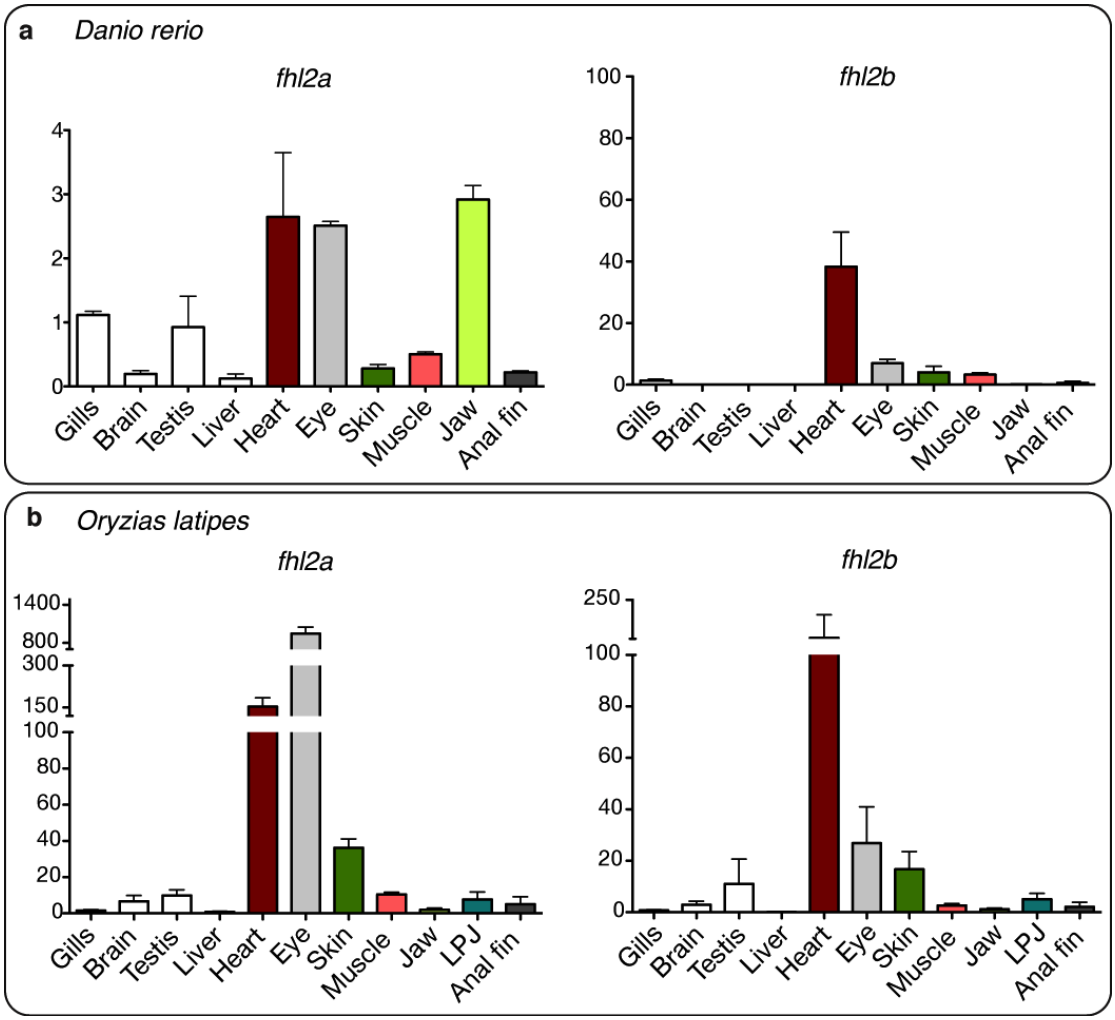


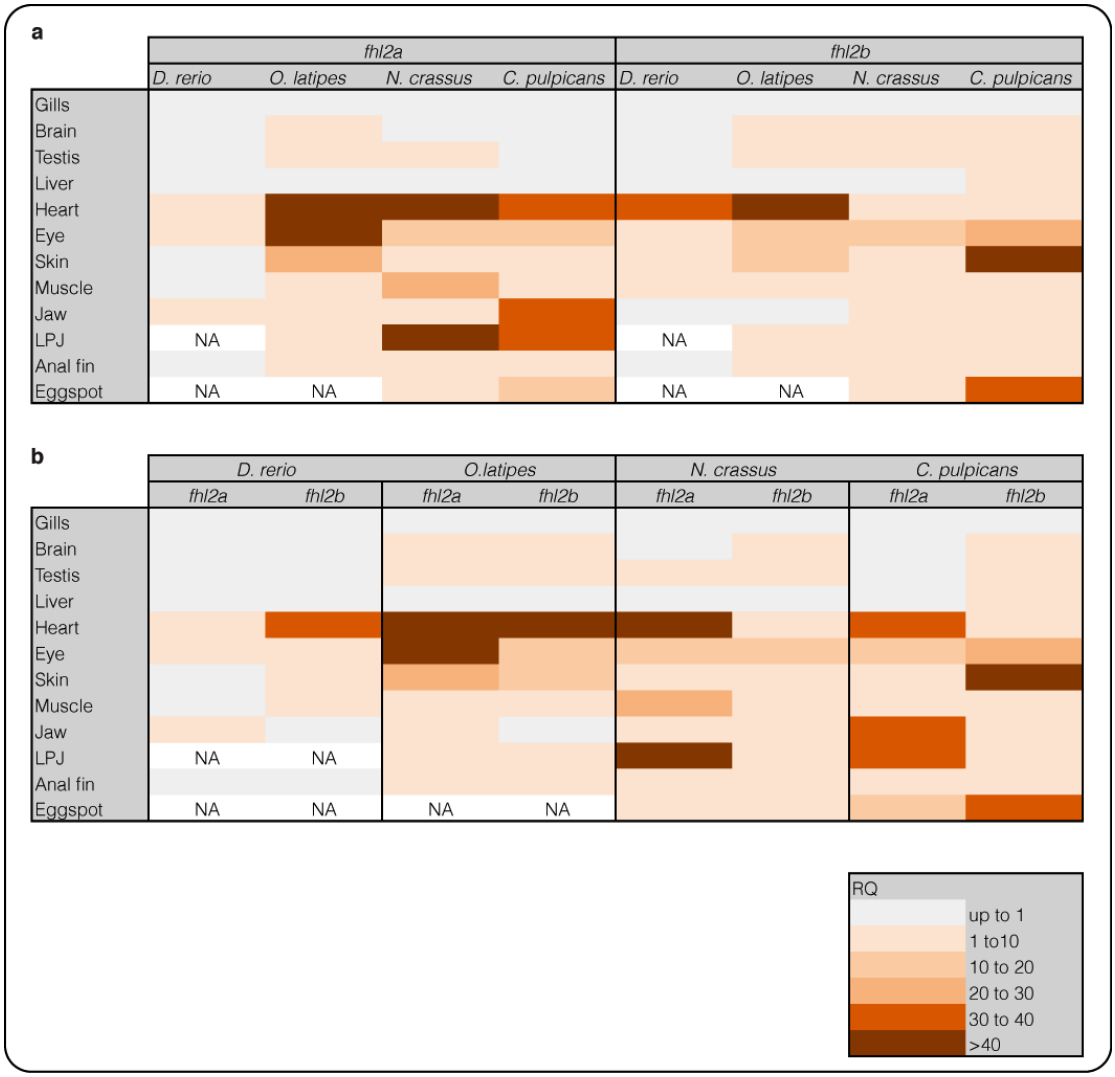
Supplementary Figure 1 | Phylogeny of East African cichlids based on a new multi-marker dataset. (**a**) Bayesian inference phylogeny with MrBayes. (**b**) Maximum likelihood phylogeny with GARLI and 500 bootstrap replicates. While most of the branches are supported with high posterior probabilities (**a**) and bootstrap values (**b**), the phylogenetic relationships among the more ancestral haplochromines – including *Pseudocrenilabrus philander* – are poorly supported and differ between the analyses. LM: Lake Malawi, LV: Lake Victoria, LT: Lake Tanganyika



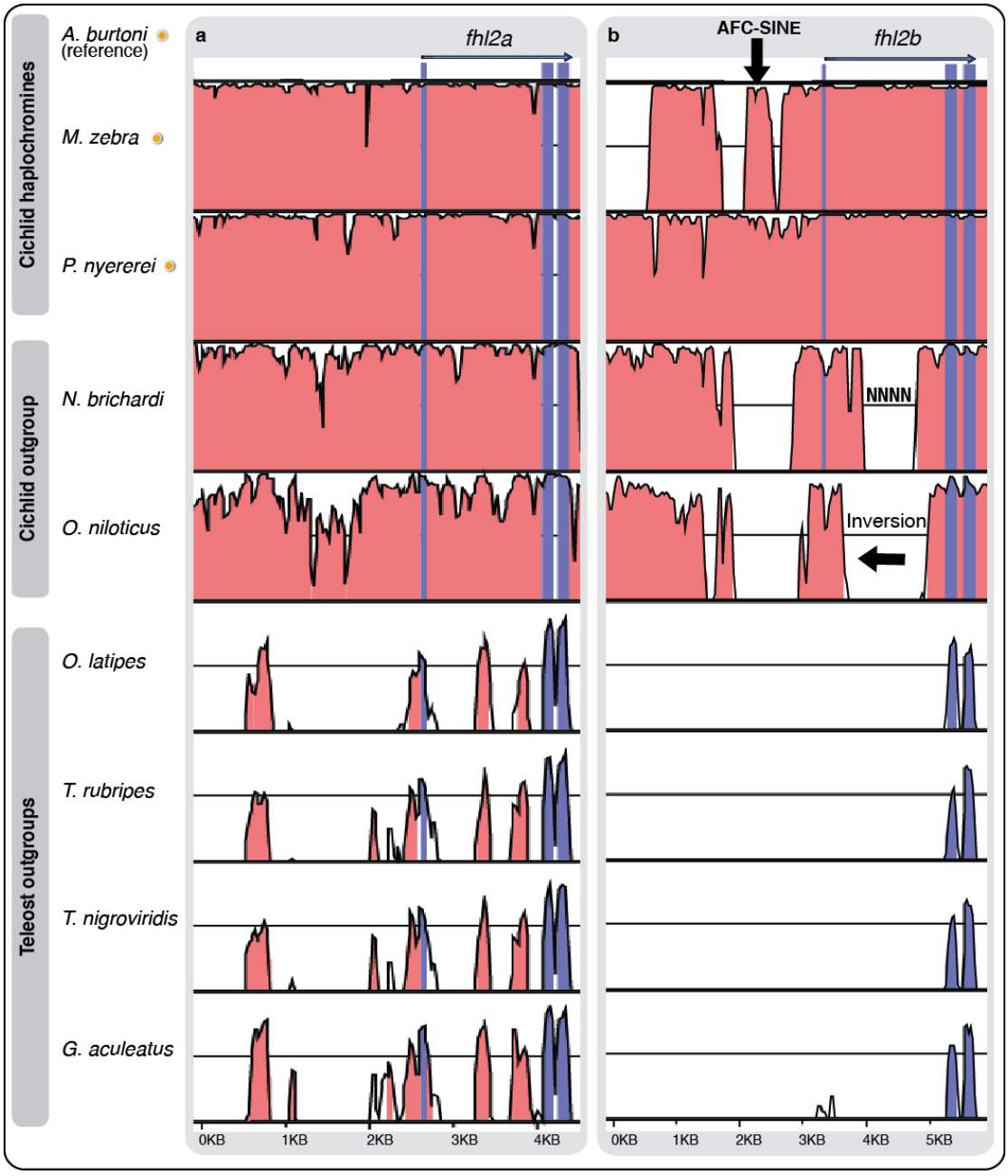
Supplementary Figure 2 | Synteny analysis of teleost *fhl2* paralogs. Dotplots of the human FHL2 gene region on human chr2 (100-220Mb) shows double conserved synteny to the two *fhl2* paralogons in (**a**) medaka on chromosomes Ola21 (*fhl2a*) and Ola2 (*fhl2b*) and in (**b**) zebrafish on chromosomes Dre9 (*fhl2b*) and Dre6 (*fhl2b*). These chromosomes were previously shown to be derived from the ancestral chromosome *c* and duplicated during the teleost genome duplication[1,2].



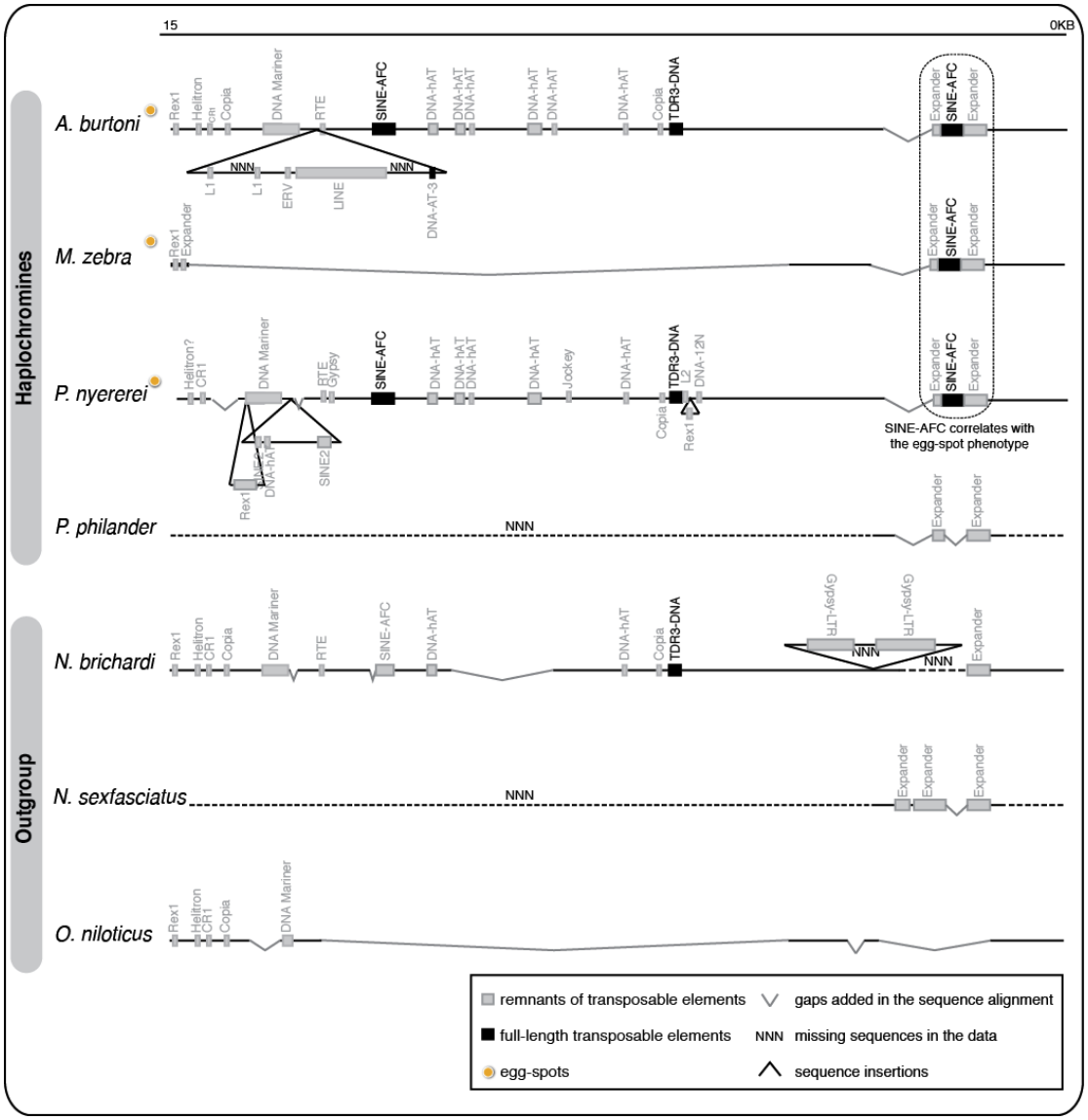
Supplementary Figure 3 | Gene expression profiling in the teleost *Danio rerio* and *Oryzias latipes.* (**a**) Relative quantification (RQ) of *fhl2a* and *fhl2b* gene expression in ten tissues in *D. rerio* (three replicates per tissue) (**b**) RQ of *fhl2a* and *fhl2b* gene expression in eleven tissues in *O. latipes* (three replicates per tissue). In both species, gill tissue was used as reference. The error bars represent the standard error of the mean (SEM). In *D. rerio* (**a**) expression of *fhl2a* is higher in heart, eye, and oral jaw, although the expression of this gene copy is overall very low, especially when compared to the level of *fhl2a* expression in cichlids and *O. latipes*. Contrary to the scenario in cichlids (Fig. 3), in *D. rerio fhl2b* is mainly expressed in the heart. In *O. latipes* (**b**) both duplicates are highly expressed in heart, skin and eye tissues. *fhl2a* does not show high expression levels in the pharyngeal jaw (unlike cichlid *fhl2a).* In this species both copies show a similar expression profile. These results suggest that the divergence history between the duplicates was different in the different lineages of teleosts, where divergence in expression profile is stronger in cichlids (see Figure 3).



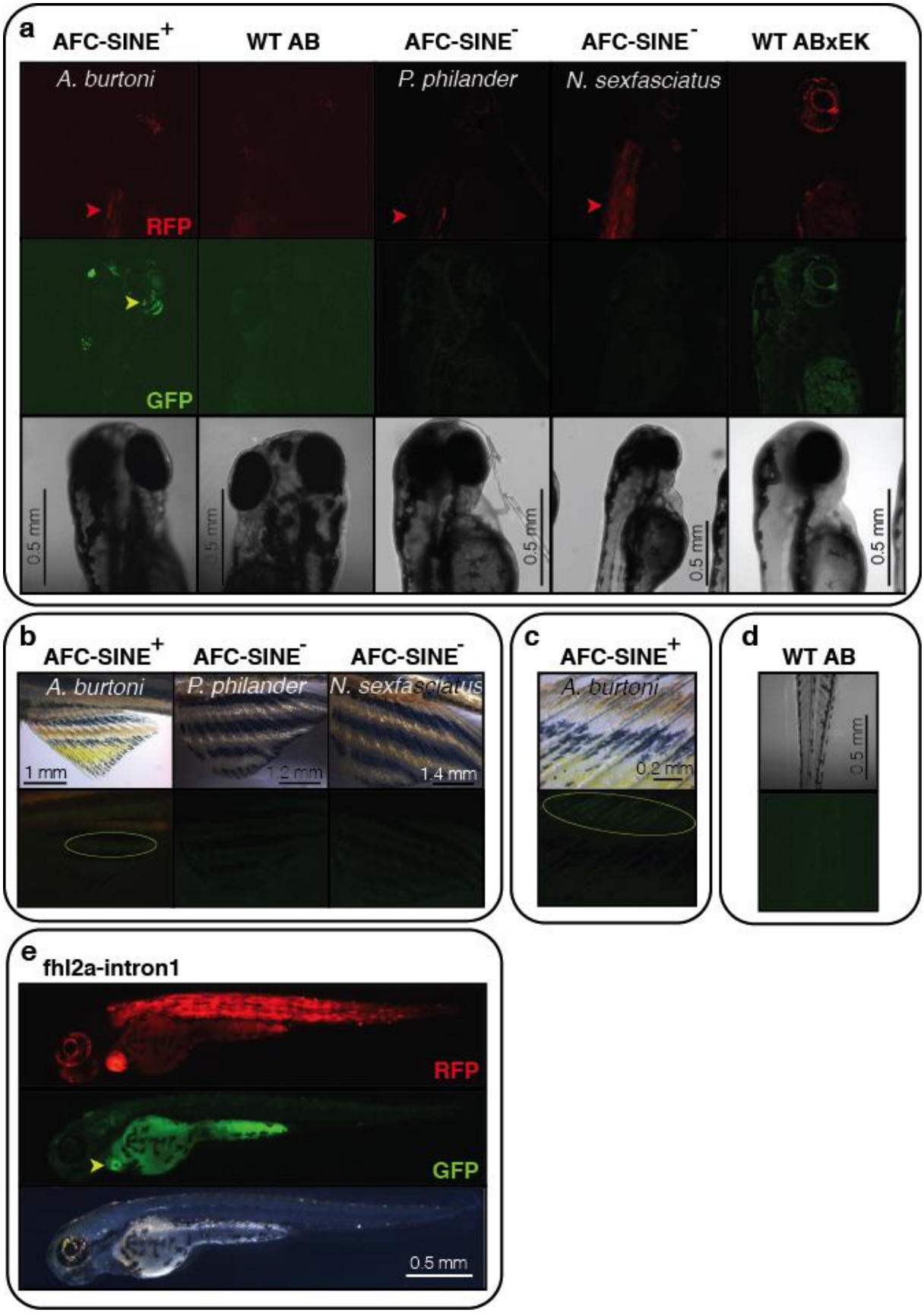
Supplementary Figure 4 | *fhl2* duplicates expression heatmap. The relative quantification values (RQ) from the qPCR experiment 3 (see methods, Fig. 3 and Supplementary Fig. 4) were categorized and color coded accordingly. Colour code and RQ value categories can be found in the bottom of this figure. The heatmap in (**a**) is grouped by gene, whereas the heatmap in (**b**) is grouped by species. The grouping of expression data by gene (**a**) suggests that both paralogs seems to have gained a new function in cichlids (*fhl2a* in jaw and LPJ and *fhl2b* in skin and egg-spot tissue). The grouping of the gene expression data by species (**b**) suggests that the expression profiles of *D. rerio* and *O. latipes* are similar, whereas the expression patterns of cichlids are more divergent.


Cichlid haplochromines
Cichlid outgroup
Teleost outgroups
A. burtoni (reference)
M. zebra
P. nyererei
N. brichardi
O. niloticus
O. latipes
T. rubripes
T. nigroviridis
G. aculeatus
a
fhl2a
b
AFC-SINE
fhl2b
NNNN
Inversion
0KB
1KB
2KB
3KB
4KB
5KB

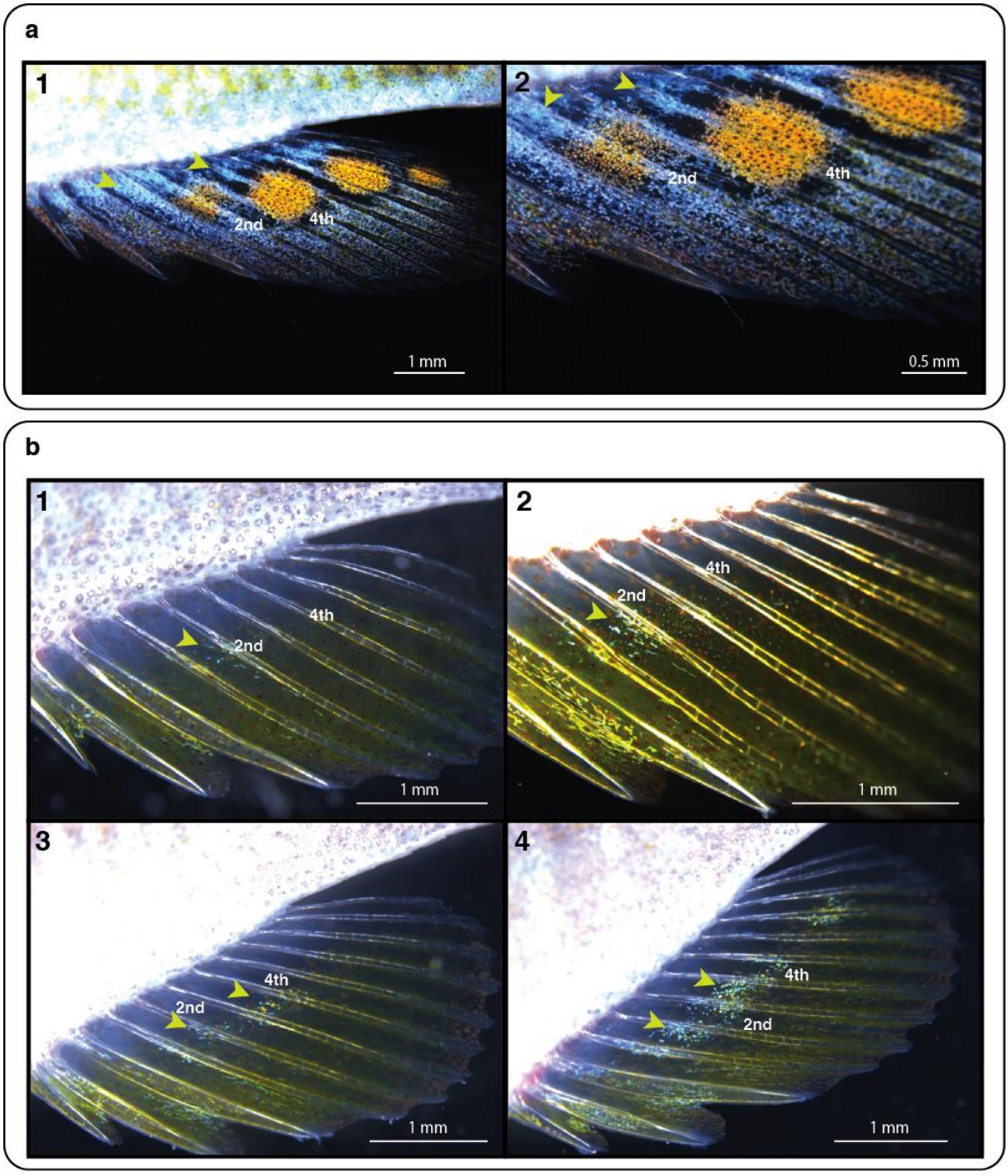
Supplementary Figure 5 | Genomic comparisons of *fhl2* paralogs of cichlids and other teleosts. *A. burtoni* genomic sequences were used as a reference for the alignments. The conserved regions (>70% identity) are marked in pink. Any conservation of the non-coding sequence across distant taxa is an indication of functional constraint and therefore of their potential role as enhancers/promoters. (**a**) MVISTA plot of a 5kb region of the *fhl2a* locus; the first three exons of *fhl2a* are shown in purple. We found four CNEs (conserved non-coding elements) and all of them are shared amongst cichlids; no major difference was detected among these CNEs. (**b**) MVISTA plot of a 5kb region of *fhl2b* locus; the first three exons of *fhl2b* are shown in purple. There are no conserved non-coding elements (CNEs) in common with the other examined teleosts, the only conserved regions are the exons. There are two regions that are conserved among haplochromines but not with non-haplochromines. One region corresponds to the first intron (positioned between 3.5-5kb), and the other region (positioned between 2-3kb) corresponds to a transposable element insertion (vertical arrow on top of the diagram), which is only present in haplochromines, the egg-spot bearing lineage. More specifically, we found that the egg-spot bearing haplochromines are characterized by an AFC-SINE insertion upstream of the *fhl2b* open reading frame. After close inspection we determined that the lack of conservation in intron one is a result of inversions (marked with a horizontal arrow) (in *O. niloticus* or in the other cichlids examined) and a missing sequence in the genome assembly of *N. brichardi* (marked with NNNN). This region most probably cannot explain the presence/absence of the egg-spot since the transcription factor binding sites would not be lost due to the inversion. These results indicate that a SINE element insertion is the likely explanation for the difference in *fhl2b* expression between haplochromines and non-haplochromines (Fig. 3).



Supplementary Figure 6 | Detailed characterization of the *fhl2b* upstream region in cichlid fishes. A more detailed characterization of the upstream genomic region of *fhl2b* shows that multiple transposon element insertions occurred in different cichlid species. It seems that this upstream region is prone to insertions (when compared to *fhl2a*), and that these insertions might disrupt the regulatory regions of this gene, explaining why *fhl2b* is more divergent in terms of gene expression than *fhl2a*. The AFC-SINE insertion is the only transposable element insertion that correlates with the egg-spot phenotype though. Interestingly, *M. zebra* (haplochromine) has several deletions in the *fhl2b* upstream region, but the AFC-SINE element is still conserved suggesting that this element is functionally important.

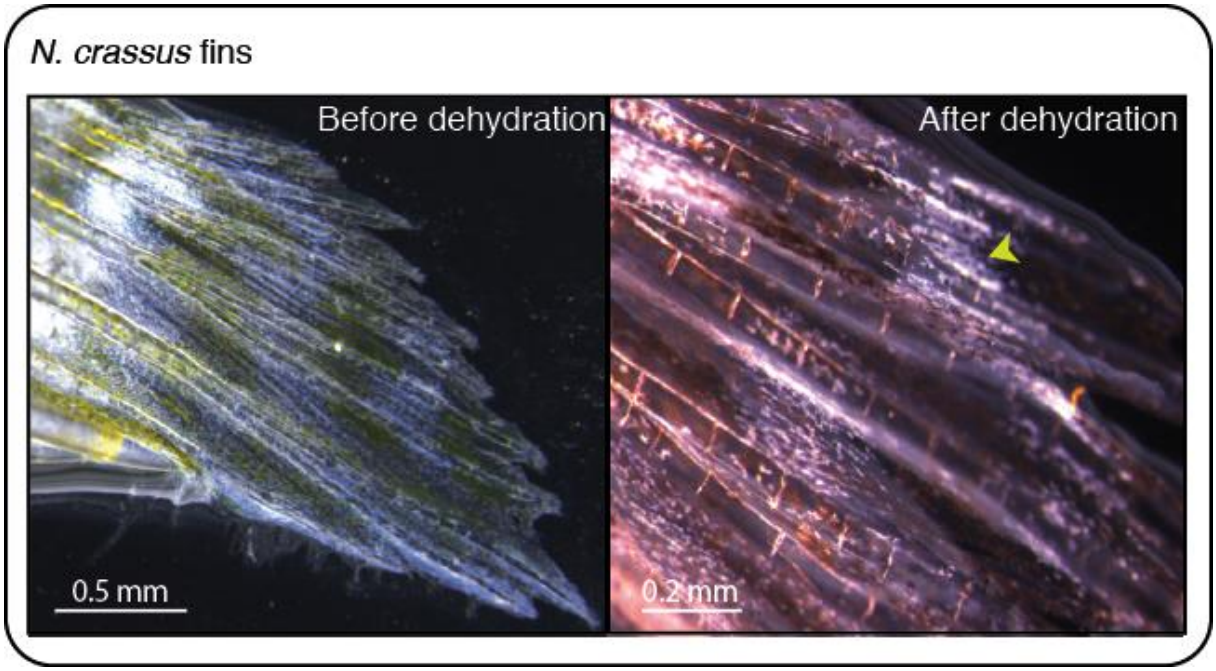

a
AFC-SINE+
WT AB
AFC-SINE-
AFC-SINE-
WT ABxEK
A. burtoni
P. philander
N. sexfasciatus
RFP
GFP
0.5 mm
b
AFC-SINE+
AFC-SINE-
AFC-SINE-
A. burtoni
P. philander
N. sexfasciatus
1 mm
1.2 mm
1.4 mm
c
AFC-SINE+
A. burtoni
0.2 mm
d
WT AB
0.5 mm
e
fhl2a-intron1
RFP
GFP
0.5 mm

Supplementary Figure 7 | Molecular and cellular basis of egg-spot development. (**a**) The upper panel shows images of the head region of a 3 day-old zebrafish embryo under RFP filter, the middle panel is the same region under GFP light and the lower panel under bright field light. The RFP shows that transgenesis was effective (positive control, for more information see ref. 3). The *A. burtoni* AFC-SINE^{+} construct drives GFP expression in the iridophores of the embryo eye (yellow arrow), a pattern that is not seen with any of the AFC-SINE^{-} constructs (*P. philander* and *N. sexfasciatus*). The two wild-type strains (AB and ABxEK) used in this study were also imaged. (**b**) Only the AFC-SINE^{+} construct drives GFP expression in the iridophores of the adult anal fin in zebrafish (the stripe of iridophores is surrounded by yellow circle). (**c**) Higher magnification image of the anal fin of the AFC-SINE^{+} construct showing GFP expression in the iridophores. (**d**) Zebrafish wild-type strain AB. This image complements the main manuscript figure 4 where we only show the imaging for the wild-type strain ABxEK. (**e**) The construct containing the first exon and intron of *fhl2a* of *A. burtoni* drove expression in heart in zebrafish. Note however, that this experiment is not exactly comparable to the one with *fhl2b*, as the fhl2a construct did not contain the upstream region.

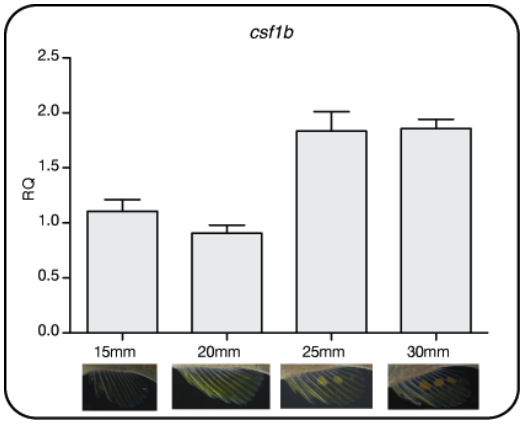


Supplementary Figure 8 | Iridophores in *A. burtoni* adult and developing fins.
(a) The upper panels show an anal fin of a juvenile *A. burtoni*. Panel 2 is a close up of panel 1. In addition to the high density of iridophores in the egg-spots (as shown in Figure 4e) we also find iridophores in high density in the proximal region of the fin (depicted with yellow arrows), which coincides with *fhl2a/b* expression patterns (see Figure 2c for *fhl2b*), suggesting that these genes are indeed iridophore genes. The first and second egg-spot are located in the second and fourth soft fin ray, respectively. Interestingly, iridophores are present in the fin rays, but absent from the inter-rays around the egg-spots, hinting at potential inhibitory relationships between different cellular types. Female anal fins (not shown) of *A. burtoni* show much smaller spots, which, in addition, emerge at later stages, and contain a much smaller number of xanthophores and iridophores. **(b)** The lower panels show different

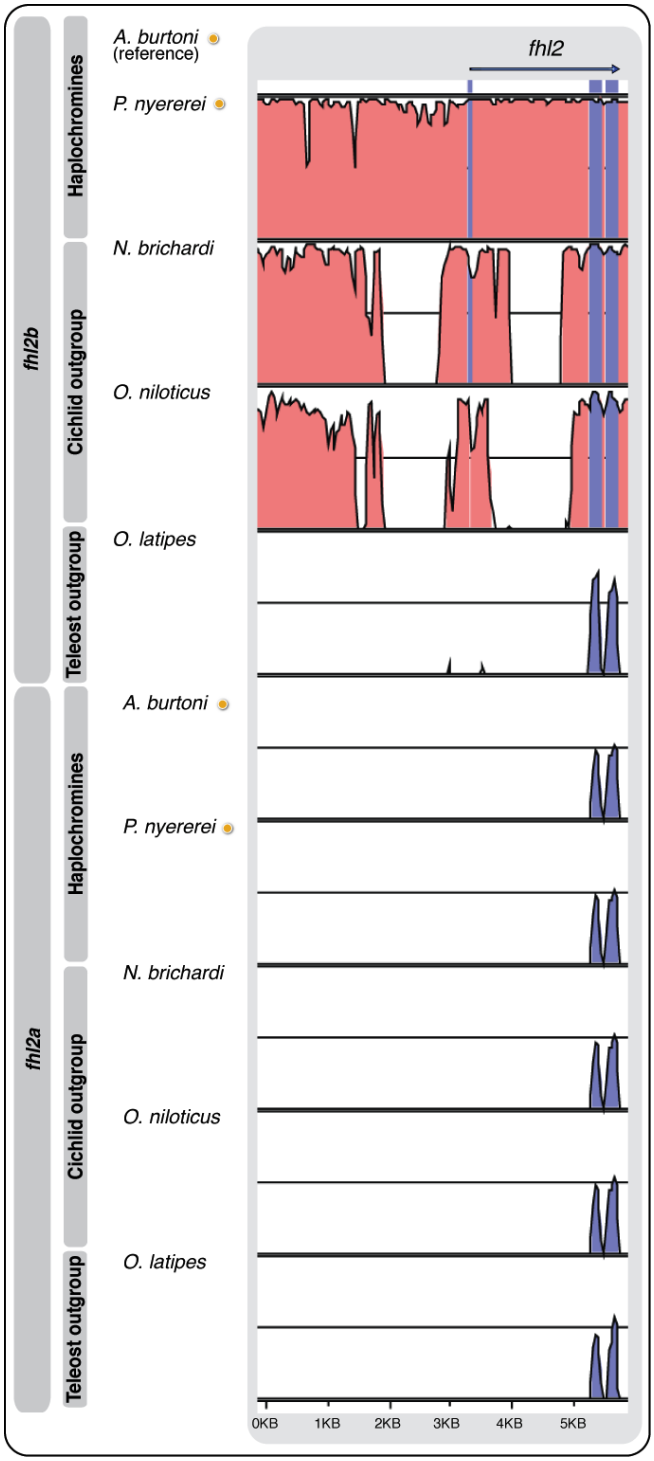
early stages of egg-spot development. The iridophores (blue cells depicted with yellow arrows) start accumulating in the second soft fin ray and then on the fourth (panel 1 and 2). At later stages (panel 3 and 4) xanthophores (yellow cells) start to accumulate where iridophores are present.



Supplementary Figure 9 | Characterization of pigment cells present in the non-haplochromine *N. crassus* fin. To better understand the morphological differences between haplochromine and non-haplochromine fins we dissected *N. crassus* anal fins. The fins of *N. crassus* show a yellow and silver transparent banded pattern (see left panel). The yellow bands are composed of xanthophores that are not visible after dehydration as the pteridine pigments are washed away. Contrary to fins with egg-spots, there are no iridophores in the xanthophore-rich regions. Instead the iridophores concentrate in the silver transparent band. Examples of iridophores are highlighted with a yellow arrow.



Supplementary Figure 10 | Expression of *csf1b* in egg-spot development. In addition to the expression profiles of *fhl2* duplicates (Fig. 2), *csf1ra*, *mitfa* and *pnp4a*, we studied the expression profile of *csf1b* (a key gene in mediating iridophores-xanthophore interaction in zebrafish) during the development of egg-spots in *A. burtoni*. The values on the x-axis represent fish standard length in millimeters. *csf1b* doubles in expression when the egg-spots emerge.



Supplementary Figure 11 | Genomic comparisons between *fhl2a* and *fhl2b* genomic region. The *A. burtoni fhl2b* genomic sequence was used as a reference for the alignment. As in Supplementary Fig. 3 the conserved regions (>70% identity) are marked in pink and conservation of the non-coding sequence across these two genes could be an indication of shared enhancers/promoters. There are no CNEs (conserved non-coding elements) shared between the two paralogs. This comparison shows that that the two regulatory regions are different.

Supplementary Table 1| Multimarker dataset used for phylogenetic analyses. The table lists GenBank accession numbers and the models of molecular evolution used in the analyses.

	Nuclear marker									
Species	*rag*	*gapdhs*	*s7*	*bmp4*	*ednrb1*	*mitfa*	*tyr*	*hag*	*csfr1a*	*nd2*
Astatotilapia burtoni	KM263618	KM263633	KM263648	KM263663	KM263678	KM263693	KM263708	KM263723	KM263738	AY930060
Astatoreochromis alluaudi	KM263630	KM263645	KM263660	KM263675	KM263690	KM263705	KM263720	KM263735	KM263750	JN628855
Pundamilia nyererei	KM263628	KM263643	KM263658	KM263668	KM263688	KM263703	KM263718	KM263733	KM263748	-
Metriaclima zebra	AGTA02035876	AGTA02043695	AGTA02046533	AGTA02003990	AGTA02044321	AGTA02044722	AGTA02007920	AGTA02058917	AGTA02053932	GU192142
Cynotilapia pulpican	KM263629	KM263644	KM263659	KM263667	KM263689	KM263704	KM263719	KM263734	KM263749	-
Ctenochromis horei	KM263623	KM263638	KM263653	KM263673	KM263683	KM263698	KM263713	KM263728	KM263743	JQ755337
Lobochilotes labiatus	KM263624	KM263639	KM263654	KM263671	KM263684	KM263699	KM263714	KM263729	KM263744	EF679250
Gnathochromis permaxillaris	KM263626	KM263641	KM263656	KM263672	KM263686	KM263701	KM263716	KM263731	KM263746	AY682522
Pseudocrenilabrus philander	KM263622	KM263637	KM263652	KM263674	KM263682	KM263697	KM263712	KM263727	KM263742	JX910862
Callochromis macrops	KM263620	KM263635	KM263650	KM263665	KM263680	KM263695	KM263710	KM263725	KM263740	CMU07242
Ophthalmotilapia ventralis	KM263621	KM263636	KM263651	KM263666	KM263681	KM263696	KM263711	KM263726	KM263741	AY337774
Aulonocranus dewindtii	KM263619	KM263634	KM263649	KM263664	KM263679	KM263694	KM263709	KM263724	KM263739	AY337782
Julidochromis ornatus	KM263627	KM263642	KM263657	KM263669	KM263687	KM263702	KM263717	KM263732	KM263747	EF462229
Limnochromis abeelei	KM263625	KM263640	KM263655	KM263670	KM263685	KM263700	KM263715	KM263730	KM263745	AY682535
Oreochromis niloticus	AERX01005120	AERX01010358	AERX01016503	AERX01008958	AERX01030771	AERX01008591	AERX01007002	AERX01042180	AERX01036679	GU477624
Neolamprologus brichardi	AFNY01022509	AFNY01036835	AFNY01033177	AFNY01048459	JF900291	AFNY01012153	AFNY01063785	AFNY01102376	AFNY01005183	AP006014
Thoracochromis brauschi	KM263632	KM263647	KM263662	KM263677	KM263692	KM263707	KM263722	KM263737	KM263752	AY930095
Serranochromis macrocephalus	KM263631	KM263646	KM263661	KM263676	KM263691	KM263706	KM263721	KM263736	KM263751	EF393690
sequenced basepairs	418	462	487	468	440	425	513	427	366	1045
used model in Garli/MrBayes (BIC)	JC	JC	HKY	HKY	HKY	HKY	HKY	K80	K80+G	TrN+G /GTR +G

Supplementary Table 2 | Top 10 differential expressed transcripts between female and male anal fins of *A. burtoni* and their identification as determined by BLASTx searches against the NCBI non-redundant database[4]. From these ten transcripts, three were identified as *fhl2* - four and a half LIM domain protein 2 - and these were among the most differentially expressed genes. Whilst tetrapods have one copy of *fhl2*, the majority of teleosts have two copies due to the extra whole genome duplication[5]. We aligned the three transcripts and observed that they are, instead of one gene, two duplicates - *fhl2a* and *fhl2b* (identification via protein homology with other teleosts and by phylogenetic inference). Comp19010_c0_seq1 and comp17680_c0_seq1 correspond to two different parts of *fhl2b* gene and comp2939_c0_seq1 corresponds to *fhl2a*. logFC stands for log_2Fold-Change in gene expression between male fins and female fins. The short read sequences are deposited under the BioProject ID PRJNA257552.

Differential expressed transcripts			**BLAST Identification**		
Transcript	**logFC**	**p-value**	**Description**	**Accession**	**e-value**
comp19010_c0_seq1 (*fhl2b*)	-5.296931976	2.55E-12	PREDICTED: four and a half LIM domains protein 2-like (Oreochromis niloticus)	XP_003446591.1	2.00E-119
comp17680_c0_seq1 (*fhl2b*)	-5.106938251	4.63E-12	PREDICTED: four and a half LIM domains protein 2-like (Oreochromis niloticus)	XP_003446591.1	3.00E-84
comp11583_c0_seq1	-4.486139111	2.41E-09	PREDICTED: similar to ORF2-encoded protein, partial (Hydra magnipapillata)	XP_002155414.1	4.00E-77
comp2939_c0_seq1 (*fhl2a*)	-4.257228806	6.57E-09	PREDICTED: four and a half LIM domains protein 2-like (Oreochromis niloticus)	XP_003453001.1	0
comp35399_c0_seq1	-4.52411029	1.32E-08	PREDICTED: crystal protein-like (Danio rerio)	XP_002661384.2	5.00E-164
comp6540_c0_seq1	-4.09287457	3.01E-08	PREDICTED: apolipoprotein D-like (Oreochromis niloticus)	XP_003448594.1	2.00E-123
comp7947_c0_seq3	-4.108859622	7.42E-08	PREDICTED: vitronectin-like (Oreochromis niloticus)	XP_003458657.1	0
comp7947_c0_seq1	-3.902264412	1.79E-07	PREDICTED: vitronectin-like (Oreochromis niloticus)	XP_003458657.1	0
comp51734_c0_seq1	-4.539615034	1.79E-07	No significant similarity found	---NA---	NA
comp7947_c0_seq2	-3.875437404	2.10E-07	PREDICTED: vitronectin-like (Oreochromis niloticus)	XP_003458657.1	0

Supplementary Table 3| Subset of primers used in this study.

Primer	Sequence 5'-3'	TºC	Task
Fhl2a_qpcr_fw	AAC ACC AGG GAT CTT TCC TAC AAG	58	qPCR in cichlids
Fhl2a_qpcr_rev	GCA CTG GAA GCA CTT AAA GCA TT	58	qPCR
Fhl2b_qpcr_fw	AGC AAG GAT CTG TCG TAC AAG GA	58	qPCR
Fhl2b_qpcr_rev	AGA CCG GCT GCA CTT GTT G	58	qPCR
RPL7_qpcr_fw	GGA GAA GTC CCT CGG CAA AT	58	qPCR
RPL7_qpcr_rev	GGC GGG CTT GAA GTT CTT TC	58	qPCR
RsPA3_qpcr_fw	AGA CCA ATG ACC TGA AGG AAG TG	58	qPCR
RsPA3_qpcr_rev	TCT CGA TGT CCT TGC CAA CA	58	qPCR
Fhl2b_probe_fw	GGT CCT CGA CTG CTA CCA AG	64	*In situ* probe PCR
Fhl2b_probe_rev	TTG CAG TTG AAG CAA TCG TT	64	*In situ* probe PCR
Fhl2a_probe_fw	CAG ACG TCC TCA GAC AGG AA	64	*In situ* probe PCR
Fhl2a_probe_rev	TGC ATC GTT CCC TGA TCA TA	64	*In situ* probe PCR
Fhl2b_TEregion_fw	CTA CTG GTG TTG GCC AGA GG	62	AFC-SINE + intron1
Fhl2b_exon2_rev	GAG AAT AGC GTC TCA TAG CAC T	62	AFC-SINE + intron1
Mitfa_fw	GCC TCG CCA TCA ACA GTT GT		qPCR in cichlids
Mitfa_rev	TCA TGC CAG GAG CAG TGA ATT		qPCR in cichlids
Csf1ra_fw	CTC AGG GCC TCG ACT TTT TG		qPCR in cichlids
Csf1ra_rev	TTC CTC GCA GCC ACA TCT C		qPCR in cichlids
Pnp4a_fw	CAT GAC CCT GGA CTG TGC TC		qPCR in cichlids
Pnp4a_rev	CTG GCT GAT GTC CCA AAC AA		qPCR in cichlids
Csf1b_fw	CCC ATG CAG ACA CTC CAT CA		qPCR in cichlids
Csf1b_rev	TTT GCT CAA ACT CCT CCG TTC		qPCR in cichlids
Ltk_fw	CTC AGG ACA GTG CTG CCA AC		qPCR in cichlids
Ltk_rev	CAG GAT GGA TCC TCC CAA AG		qPCR in cichlids
Fhl2a_Drerio_qpcr_fw	CGG CTG CGC AGA AGT AAA G		qPCR in *D. rerio*
Fhl2a_Drerio_qpcr_rev	GTA TGG GTT GTC CTC ACG CA		qPCR in *D. rerio*
Fhl2b_Drerio_qpcr_fw	CAC GGG ACA GGG ATT GTT TA		qPCR in *D. rerio*
Fhl2b_Drerio_qpcr_rev	CCC GAA CAG AGA CTC CTT GC		qPCR in *D. rerio*
Fhl2a_Olatipes_qpcr_fw	CAC TGC AAG AAG CCC ATC AC		qPCR in *O. latipes*
Fhl2a_Olatipes_qpcr_rev	ACG AAG CAC TCT TTG TGC CA		qPCR in *O. latipes*
Fhl2b_Olatipes_qpcr_fw	CCA ACA CCT GTG AGG AAT GC		qPCR in *O. latipes*
Fhl2b_Olatipes_qpcr_rev	TCA TGC CAG TGA CGG TCT TT		qPCR in *O. latipes*

Supplementary Table 4 | Sequencing of the *fhl2a* and *fhl2b* coding region sequencing in cichlids. The table lists specimens, primer sequences and annealing temperatures (the sequences are available under the accession numbers KM263753 to KM263998).

Species name	**Lineage**	**Anal fin egg-spot**	**Nr of ind.(fhl2a/fhl2b)**
Astatoreochromis alluaudi	Haplochromini	Yes	5 / 5
Astatotilapia burtoni	Haplochromini	Yes	5 / 5
Labidochromis caeruleus	Haplochromini	Yes	5 / 4
Pseudocrenilabrus multicolor	Haplochromini	Blotch	4 / 4
Pseudotropheus elegans	Haplochromini	Yes	4 / 5
Cynotilapia pulpican	Haplochromini	Yes	4 / 5
Ctenochromis horei	Haplochromini	Yes	5 / 5
Interochromis loocki	Haplochromini	Yes	4 / 2
Petrochromis famula	Haplochromini	Yes	5 / 5
Petrochromis fasciolatus	Haplochromini	Yes	5 / 5
Petrochromis polyodon	Haplochromini	Yes	4 / 5
Pseudosimochromis curvifrons	Haplochromini	Yes	5 / 5
Simochromis diagramma	Haplochromini	Yes	5 / 5
Tropheus moori	Haplochromini	Yes	5 / 5
Julidochromis marlieri	Lamprologini	No	5 / 5
Julidochromis ornatus	Lamprologini	No	5 / 5
Neolamprologus brichardi	Lamprologini	No	4 / 4
Neolamprologus pulcher	Lamprologini	No	4 / 5
Neolamprologus sexfasciatus	Lamprologini	No	5 / 5
Cyprichromis 'jumbo'	Cyprichromini	No	5 / 5
Cyprichromis leptosoma	Cyprichromini	No	5 / 5
Aulonocranus dewindtii	Ectodini	No	5 / 5
Callochromis macrops	Ectodini	Blotch	5 / 5
Cyathopharynx furcifer	Ectodini	No	4 / 5
Ophthalmotilapia ventralis	Ectodini	No	5 / 5
Xenotilapia flavipinnis	Ectodini	No	5 / 5
Primer	**Sequence 5'-3'**	**PCR annealing TºC**	**Task**
Fhl2b_CDS_fw	GGT CCT CGA CTG CTA CCA AG	65-68	Coding region PCR/Seq.
Fhl2b_CDS_rev	ATC CGG CTC GGG TTG TCT	65-68	Coding region PCR/Seq.
Fhl2a_CDS_fw	CTG CCA CAG ACT CCA CAC AG	65-68	Coding region PCR/Seq.
Fhl2a_CDS_rev	TGC ATC GTT CCC TGA TCA TA	65-68	Coding region PCR/Seq.

Supplementary Table 5 | Species (and GenBank accession numbers) used to infer the *fhl2* gene tree.

Species name	Lineage	*fhl2a*	*fhl2b*	*fhl2*
Astatoreochromis alluaudi	Cichlidae	KM263877	KM263753	NA
Astatotilapia burtoni	Cichlidae	KM263882	KM263758	NA
Pseudocrenilabrus philander	Cichlidae	KM263976	KM263853	NA
Cynotilapia pulpican	Cichlidae	KM263980	KM263857	NA
Ctenochromis horei	Cichlidae	KM263897	KM263773	NA
Petrochromis fasciolatus	Cichlidae	KM263958	KM263833	NA
Pseudosimochromis curvifrons	Cichlidae	KM263967	KM263843	NA
Simochromis diagramma	Cichlidae	KM263984	KM263862	NA
Tropheus moori	Cichlidae	KM263989	KM263867	NA
Julidochromis ornatus	Cichlidae	KM263925	KM263800	NA
Neolamprologus brichardi	Cichlidae	KM263935	KM263809	NA
Neolamprologus sexfasciatus	Cichlidae	KM263943	KM263818	NA
Cyprichromis 'jumbo'	Cichlidae	KM263906	KM263783	NA
Cyprichromis leptosoma	Cichlidae	KM263911	KM263788	NA
Aulonocranus dewindtii	Cichlidae	KM263887	KM263763	NA
Callochromis macrops	Cichlidae	KM263892	KM263768	NA
Cyathopharynx furcifer	Cichlidae	KM263902	KM263778	NA
Xenotilapia flavipinnis	Cichlidae	KM263994	KM263872	NA
Oryzias latipes	Adrianichthyidae	ENSORLG00000012482	ENSORLG00000001848	NA
Takifugu rubripes	Tetraodontidae	ENSTRUG00000013559	ENSTRUG00000008468	NA
Gasterosteus aculeatus	Gasterosteidae	ENSGACG00000003005	ENSGACG00000015048	NA
Danio rerio	Cyprinidae	ENSDARG00000042018	ENSDARG00000003991	NA
Anolis carolinensis	Polychrotidae	NA	NA	ENSACAG00000010422
Mus musculus	Muridae	NA	NA	ENSMUSG00000008136

Supplementary Table 6 | Species (with corresponding GenBank accession numbers) used to infer the ND2 phylogeny for the positive selection analysis.

Species name	Cichlid Lineage	ND2 accession number
Astatoreochromis alluaudi	Haplochromini	AY930071.1
Astatotilapia burtoni	Haplochromini	AF317.266.1
Labidochromis caeruleus	Haplochromini	AY740383.1
Pseudocrenilabrus multicolor	Haplochromini	AY930106.1
Pseudotropheus tropheops	Haplochromini	AY740384
Cynotilapia pulpican	Haplochromini	-
Ctenochromis horei	Haplochromini	EU753935
Interochromis loocki	Haplochromini	JF900322
Petrochromis famula	Haplochromini	JF900324
Petrochromis fasciolatus	Haplochromini	JF900325
Petrochromis polyodon	Haplochromini	JF900326
Pseudosimochromis curvifrons	Haplochromini	GQ995777.1
Simochromis diagramma	Haplochromini	AY930087
Tropheus moori	Haplochromini	AY930093
Julidochromis marlieri	Lamprologini	AF398230
Julidochromis ornatus	Lamprologini	EF462229
Neolamprologus brichardi	Lamprologini	AF398227
Neolamprologus pulcher	Lamprologini	EF462244
Neolamprologus sexfasciatus	Lamprologini	HM623828
Cyprichromis 'jumbo'	Cyprichromini	AF317.266.1
Cyprichromis leptosoma	Cyprichromini	AF398224
Aulonocranus dewindtii	Ectodini	AY337782
Callochromis macrops	Ectodini	AY337795
Cyathopharynx furcifer	Ectodini	AY337781
Ophthalmotilapia ventralis	Ectodini	AY337774
Xenotilapia flavipinnis	Ectodini	AY337794
Oreochromis niloticus	Tilapini	AF317237
Oryzias latipes	non-cichlid teleost	NC_004387.1

Supplementary Table 7 | Testing for branch or site-specific positive selection on East African cichlids *fhl2a* and *fhl2b* with CODEML.

We found no evidence for branch or site-specific positive selection in the *fhl2a* copy while in *fhl2b*, we detected positive selection on three amino acids (positions 10, 86 and 150). Positive selection was only detected within one species, *Pseudotropheus elegans* (Pseele). Overall, *fhl2a* and *fhl2b* are under purifying selection, showing that coding sequences alone cannot explain the emergence or diversity of the egg-spot trait in haplochromines. All *fhl2a* and *fhl2b* coding sequences were found to be 837bp long (279 amino acids), except for three individuals from two species - *Tropheus moori* (843bp and 745bp, respectively) and *Simochromis diagramma* (840bp)

Briefly, under the null model, the foreground branch (branch of interest) has proportions of sites under neutral selection that may differ from those on the background branches. In the alternative model, positive selection is allowed on the foreground branch. For clarity's sake, we do not show the proportion of sites in each category, only the computed value of the corresponding *dn/ds* ratios (ω).
In both models and on both branches:

0: $\omega_0 < 1$

1: $\omega_1 = 1$

In the Null model:

2a: $\omega_2 = 1$ on foreground and $\omega_0 < 1$ on background.

2b: $\omega_2 = 1$ on foreground branch and $\omega_1 = 1$ on background.

In the Alternative model:

2a: $\omega_2 \geq 1$ on foreground and $\omega_0 < 1$ on background.

2b: $\omega_2 \geq 1$ on foreground and $\omega_1 = 1$ on background.

With ω_0: dn/ds < 1, ω_1: dn/ds = 1, ω_2: dn/ds ≥ 1. *LRT:* Likelihood Ratio Test computed as $2\times(\ln L_1 - \ln L_0)$ where L_1 is the Likelihood for the Alternative model and L_0 is the likelihood of the Null model. Under the Null model, the LRT follows a Chi-square distribution with 1 df. *Site:* positively selected amino-acid site with the amino acid change in brackets. *BEB:* Bayes Empirical Bayes.

Foreground Branch	Model	Background ω				Foreground ω				LRT	*P*-value	Site	BEB
	Site classes	**0**	**1**	**2a**	**2b**	**0**	**1**	**2a**	**2b**				
***fhl2a* locus**													
Haplochromini (*Psephi excluded)*	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	-3.26212	1	--	--
Haplochromini + Tropheini (*Psephi excluded)*	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	0	1	--	--
Haplochromini + Tropheini (*Psephi included)*	Null	0.00512	1.00000	0.00512	1.00000	0.00512	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	-0.548136	1	--	--
Tropheini	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	-0.245402	1	--	--
Pseele	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	3.26212	0.07089741	--	--
Cichlids (including Tilapia)	Null	0.00490	1.00000	0.00490	1.00000	0.00490	1.00000	1.00000	1.00000				
	Alternative	0.00490	1.00000	0.00490	1.00000	0.00490	1.00000	1.00000	1.00000	0.00000	1	--	--
***fhl2b* locus**													
Haplochromini (*Psephi excluded)*	Null	0.04541	1.00000	0.04541	1.00000	0.04541	1.00000	1.00000	1.00000			86 (Q)	0.685
	Alternative	0.04532	1.00000	0.04532	1.00000	0.04532	1.00000	**76.33684**	**76.33684**	**7.805352**	**0.00520917**	150 (C)	0.954*
Haplochromini + Tropheini (*Psephi excluded)*	Null	0.04471	1.00000	0.04471	1.00000	0.04471	1.00000	1.00000	1.00000				
	Alternative	0.04588	1.00000	0.04588	1.00000	0.04588	1.00000	22.94605	22.94605	3.832012	0.05028254	--	--
Haplochromini + Tropheini (*Psephi included)*	Null	0.04471	1.00000	0.04471	1.00000	0.04471	1.00000	1.00000	1.00000				
	Alternative	0.04588	1.00000	0.04588	1.00000	0.04588	1.00000	22.94743	22.94743	3.832242	0.05027564	--	--
Tropheini	Null	0.04754	1.00000	0.04754	1.00000	0.04754	1.00000	1.00000	1.00000				
	Alternative	0.04768	1.00000	0.04768	1.00000	0.04768	1.00000	1.00000	1.00000	-0.004042	1	--	--
Pseele	Null	0.04498	1.00000	0.04498	1.00000	0.04498	1.00000	1.00000	1.00000			10 (C)	0.950*
	Alternative	0.04484	1.00000	0.04484	1.00000	0.04484	1.00000	635.39848	**635.39848**	**12.508918**	**0.00040501**	86 (L)	0.593
												150 (C)	0.998**
Cichlids (including Tilapia)	Null	0.03499	1.00000	0.03499	1.00000	0.03499	1.00000	1.00000	1.00000				
	Alternative	0.03998	1.00000	0.03998	1.00000	0.03998	1.00000	8.89460	8.89460	1.364914	0.242688	--	--

Supplementary Table 8 | The AFC-SINE insertion is specific to egg-spot bearing haplochromines. In order to test whether the SINE insertion is correlated with the egg-spot phenotype we sequenced this transposable element region in 19 cichlid species, including both haplochromines and non-haplochromines. We also sequenced one haplochromine species that has no egg-spot but instead features a blotch on its anal fin: *Pseudocrenilabrus philander*. We confirmed that the AFC-SINE insertion is specific to the egg-spot bearing haplochromines, whereas the ancestral haplochromine *P. philander* does not have this insertion. The AFC-SINE element in the *fhl2b* promoter region was compared to the consensus sequence and available full-length AFC-SINE elements of cichlids. The insertion in the *fhl2b* promoter covers a full-length element and is flanked both 3' and 5' by five nucleotide long direct repeats, so called target site duplications. These duplications are the result of the element's insertion process via DNA strand break and repair[6], confirming that this is an insertion in haplochromines and not a deletion in non-haplochromines, since no remnants of these sites or the element are detected in the other species. The sequences are available under the accession numbers KM263999 to KM264016.

Species Name	**Lineage**	**Anal Fin Eggspot**	**AFC-SINE**
Astatotilapia burtoni	Haplochromini	Yes	Present
Astatotilapia stapersi	Haplochromini	Yes	Present
Astatoreochromis alluaudi	Haplochromini	Yes	Present
Pundamilia nyererei	Haplochromini	Yes	Present
Metriaclima zebra	Haplochromini	Yes	Present
Cynotilapia pulpican	Haplochromini	Yes	Present
Ctenochromis horei	Tropheini	Yes	Present
Lobochilotes labiatus	Tropheini	Yes	Present
Gnathochromis pfefferi	Tropheini	Yes	Present
Limnotilapia dardenii	Tropheini	Yes	Present
Pseudocrenilabus philander	Basal Haplochromini	No	Absent
Callochromis macrops	Ectodini	No	Absent
Ophthalmotilapia ventralis	Ectodini	No	Absent
Ophthalmotilatpia nasuta	Ectodini	No	Absent
Aulonocranus dewindtii	Ectodini	No	Absent
Perissodus microlepis	Perissodini	No	Absent
Neolamprologus sexfasciatus	Lamprologini	No	Absent
Julidochromis ornatus	Lamprologini	No	Absent
Lamprologus lemairii	Lamprologini	No	Absent
Limnochromis abeeli	Limnochromini	No	Absent
*Oreochromis niloticus**	Tilapini	No	Absent
Primer	**Sequence 5'-3'**	**TºC**	**Task**
Fhl2b_TEregion_fw	GAAGTCATGCAATGACAGACA	58-60	PCR TE region
Fhl2b_TEregion_rev	AATCCTCTGGGCAAAATGTGC	58-60	PCR TE region
Fhl2b_Hap_fw **	CTACTGGTGTTGGCCAGAGG	59-60	Seq TE region
Fhl2b_NonHap_fw **	TTAAAGTCATTAAAATGTCCCGATT	59-60	Seq TE region

*only available from genome not amplified in house

**primers used only for sequencing and not for product amplification, Hap only works for haplochromines and non-Hap only works for non-haplochromines

Supplemental References

1. Kasahara, M. *et al.* The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447,** 714–9 (2007).

2. Nakatani, Y., Takeda, H., Kohara, Y. & Morishita, S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* **17,** 1254–65 (2007).

3. Bessa, J. *et al.* Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev. Dyn.* **238,** 2409–2417 (2009).

4. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33,** D501–4 (2005).

5. Meyer, A. & Van de Peer, Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* **27,** 937–945 (2005).

6. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).